Replies to Referee #1 on the manuscript 'Assessment of the error budget for stratospheric ozone profiles retrieved from OMPS limb-scatter measurements ' by C. Arosio et al.

We thank the reviewer for the time she/he spent reading the manuscript and constructively commenting on the paper. In the text below, we address the comments from the Referee #1. Referee's comments are shown in italic and authors' responses are highlighted in blue.

—— General comments

This is a nice solid paper that does a thorough and robust job of quantifying the uncertainties associated with the Bremen group's retrievals of ozone profiles from the OMPS-LP instrument. My comments are only minor. The discussion is generally very clear and well motivated. The language and graphics are both of very good quality. I'm very happy to recommend this paper for publication pending the minor corrections that I detail below. I do not feel a need to review an updated version (but clicked on the "willing" button if the editor feels another read through is needed).

Thank you for the appreciation and the positive review, we address the comments below.

—— Specific comments

Line 2: I think the mission is now called the "Suomi National Polar-orbiting ..., Suomi NPP", right?

True, we changed the text accordingly and use now the SNPP acronym.

Line 3: I suggest rewording as: 'The error characteristics are presented in a form that is compliant with ...'. Note the spelling error on compliant (you have "complaint").

Thanks, we re-worded as 'The error characteristics are presented in a form that aims at being compliant with...'

Line 8: Suggest rewording start of sentence to: "We have classified the various different error..."

Done.

Line 11: You have a short hyphen (single "-" character in LaTeX) for the first (3-50%) range, but longer ones ("–" in LaTeX, which are more correct) for the other two. Modify the first to be consistent/correct.There were some other examples of this issue elsewhere in the paper.

Thanks for pointing this out, we replaced the short hyphens with long ones in those cases.

Line 12: "related" → "due"

Done.

Line 13: saying both "absolute value" and quoting a percent (which many would see as a "relative value" is confusing). Please clarify, do you mean abs(error) - to use programming terminology? - expressed as a percent?

Yes, exactly, we meant errors in 'absolute value', we removed the word absolute, which can lead to confusion.

Line 13: What above between 20 and 50 km, would be good to quantify that in abstract too.

We changed the sentence to '... The corresponding total bias exceeds 5 % only above 50 km and below 20 km'.

Line 18: Insert "Suomi" before "National"

Done.

Line 26: "was in the meantime" → "has since been"

Done.

Line 41: Put commas before and after "as in von Clarmann et al. (2020)"

Done.

Line 43: Would it be better to insert "mainly" before "contribute" here? I leave it to you.

Done.

Line 55: Suggest inserting "in Sect. 2.1" after "instrument" for better flow/consistency. Then insert "of" before "the retrieval algorithm" and delete "given"

Done.

Line 70: Suggest "increases" → "coarsens". "Increased" resolution sounds good from an English perspective, but is worse from a numerical perspective if quantified in terms of a length. (Similarly precision/accuracy. I generally advocate saying coarser/finer and/or better/worse rather than more/less for all such terms to avoid ambiguity).

Thanks, it is a good point, we used 'degrades' instead.

Line 81: Delete "what concerns"

Done, we replaced it with 'to choose'.

Line 86: "... and are preconvolved with the OMPS-LP instrument line shape" ("are beforehand" is awkward wording).

Done.

Line 90 (and 95): Not being familiar with UV/Vis instruments, I hadn't understood/remembered that the Chappuis bands are not in the UV. Thus this paragraph was very confusing to me (is the shift and squeeze included or not?). I suggest you make that clearer by pointing out that the Chappuis bands are in the visible spectrum).

Thanks, we didn't realize this. We clarified this in the text.

Line 91: "tells" → "suggests" or "implies"

Replaced with 'implies'.

Line 94: Not sure what is meant by "differential absorption" here, I presume you mean fine-scale spectral structure, but it's not 100% obvious. Saying "differential" could mean from TH to TH. Please clarify.

Thanks, yes we meant fine-scale spectral structures, so we changed 'differential' to 'fine-scale'.

Line 109: Suggest you delete "percentage" and then add ", expressed as a percentage of the true profile," after "reported" on the next line.

Done.

Line 118: Not clear what "actual" means in this context. Does it refer to the spacing of the levels in the state vector? (I forget if that was discussed anywhere in the paper). If so, then I'm not quite sure how the UV vs. Chappuis comes into it, so perhaps that's a bad guess on my part. Please clarify.

Yes, the observations are reported on a 1-km grid but the vertical resolution of the retrieved profile is coarser. We removed the word 'actual'.

Line 131: Some inconsistency here, this discussion of the Gaussian generator and the multiplication by the SNR discussed in line 133 suggest that these "noise sequences" are Normal(mu=0, sigma=1). However, the caption for figure 3 states that the lower panel is showing the "noise sequences" too, and these clearly have sigma smaller than 1.0. Not a big deal, I'm sure everything's been Done. correctly, but make the terminology consistent between the text and the figure caption.

Yes, the generated gaussian samples have sigma = 1 but they are multiplied by the sqrt(RMS). We specified it in the caption as well.

Line 132: "Simulated spectrum" may be unclear here, as we're stil dealing with real OMPS data, right, not the radiances corresponding to a modeled truth. Perhaps "fitted" might be better for this particular sentence, rather than "simulated"?

Thanks, yes we meant SCIATRAN spectrum, i.e. the one that is fitted. We specify 'fitted model spectrum'

Line 134: Suggest you say 50 rather than N here, for consistency with earlier in the sentence (some may forget that N==50).

Done.

Line 136: "profile" → "profiles" or "a first guess profile"?

We replaced with 'profiles'.

Line 137: Clarify "unperturbed" here. It's clear to me that you mean you have yet to do any perturbations associated with parameters, calibration, spectroscopy etc., but readers less familiar may be confused as your 50 noise sequences are themselves perturbations.

We see the point, we specified as follows: 'all 50 retrieved unperturbed profiles (with different noise sequences applied)...'.

Line 143: insert "is" between "what" and "ideally"

Done.

Line 143/144: "This feature" is vague - is it the point discussed in the preceding sentence or the whole of this paragraph. Please clarify.

We replaced 'features' with 'discrepancies', which should be more clear.

Line 156: Suggest you change "shift-squeeze" in the math symbol to "shift&squeeze" to avoid confusion with a minus sign (& in LaTeX). Also, the kerning for this and other things (e.g., "param", "fix", etc.) is wrong because you're in math mode. Suggest you use, e.g., $\sigma_{x,shift\&squeeze}$ *(needs amsmath.sty) or,* $\sigma_{x,shift\&squeeze}$ *if you prefer. And similar for others.*

Done, thanks, we use the \text{} option for the subscripts.

Line 166 / Table 2: For completeness/clarify you should state what these numbers convey. Are they ranges? 95% confidence intervals? 1-sigma (which is what I assume the are, from the contex), 2-sigma? This then impacts the interpretation of the resulting O3 uncertainties.

We can assume these values to be the standard deviation of the parameters, i.e. 1 σ. We added this in the paper.

Line 168. The sentence here feels like it belongs at (or near) the end of the paragraph rather than at the start. Where it is now, it interrupts the flow of discussing the input parameter uncertainties.

We moved it to the end of the paragraph.

Line 186: Suggest adding "radiance" before "noise" to remind people what you mean here.

Thanks, we expanded as '...the Gaussian noise is added to the simulated radiances...'

Line 187: I wonder if the superscript "n" might not be better as an "i"? In any case, it should be defined.

We added at the end of sentence 'where n is the running number of the noise sequence'.

Equation 7: I can sort of see why you call this a sigma rather than a delta, but it feels like it should be the latter from the definition, so take some time to explain to the reader why you're not labeling it as a delta. Indeed, it might be easier to explain if you talk about the Monte Carlo approach first rather than second. Then this delta-like parameter becomes an estimate of the

associated sigma from the MC approach.

As long as the variables are properly defined we believe there is some freedom in using the notations and there is no need to justify the used symbol. We note that the used notations are common in the ozone profile retrieval community. However, we see the reviewer's point and her/his remark before the discussion phase and decided to indicate with sigma the random uncertainties and with a delta the systematic components.

Line 214: "are going to be" → "will be". Also, end of line change "fix" to "fixed".

Done.

Line 215: "... and conveys information on the relationship between the sign of the parameter perturbation and the sign of the consequent ozone perturbation."

Thanks, implemented.

Line 224: Would be good (here or somewhere else) to point out that you're changing Temperature at all heights by 2K, not just one height (or each height in turn, right?). Indeed there are multiple temperature, pressure, TH, and aerosol perturbations that could be made, but only one albedo one, correct? (Though hydrostatic balance, if you're imposing that, reduces the degrees of freedom for Temperature/pressure/TH perturbations) Perhaps discuss this more, and explain why you didn't consider more complex (i.e., height-dependent perturbations).

Thanks, for pointing this out, it's an interesting discussion. Temperature, pressure and aerosol errors could indeed be assumed to vary randomly in altitude; though for TH it is rather a single value for all altitudes as it is related to the pointing of the instrument as a whole. We investigated the difference between considering the perturbation constant over altitude (correlated) or free to change randomly (uncorrelated). Fig. 2 shows the comparison between these two cases, in terms of ozone errors, i.e. standard deviations of 50 retrieved profiles with perturbed aerosol and temperature profiles, with correlated (magenta shaded area) and uncorrelated (green) errors. In the first case we used 50 generated noise sequences as SNR, whereas in the second we kept the SNR from the actual OMPS measurement. Superimposed are also shown the two cases of fixed perturbation, using the SNR from the OMPS observation. Differences are negligible for aerosol but not for temperature (and pressure, here not shown). We have to note that from the nature of P and T data, i.e. model, a systematic error in the altitude domain is more probable. We added a sentence in this respect at the end of Section 4 and these figures in the Supplements.

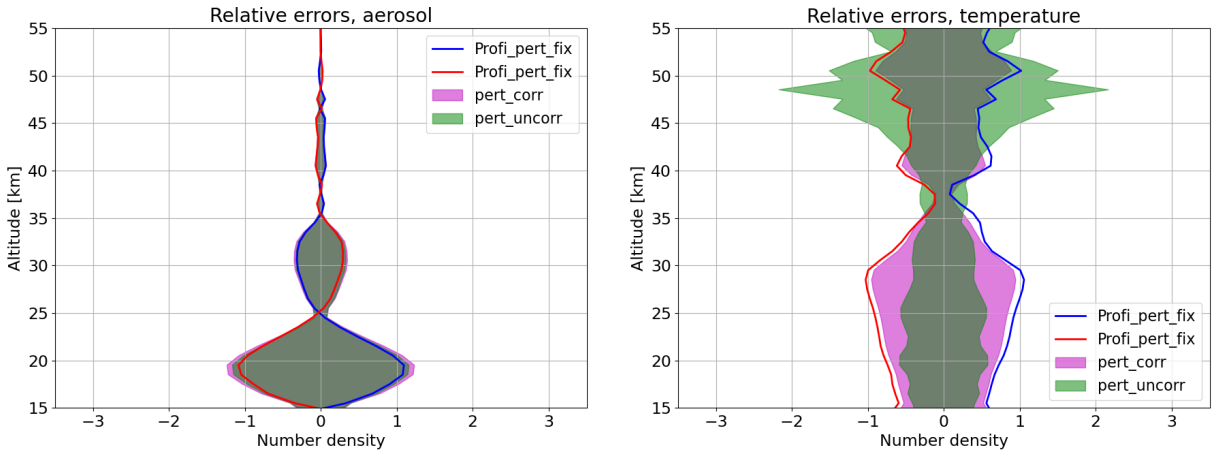

Figure 1: Comparison of estimated ozone errors when using correlated or uncorrelated parameter errors in altitude, for aerosol and temperature.

Figure 5 caption, line 2, insert "in the top and bottom rows, " before "respectively". Also, delete "- like" after "Monte Carlo"

Done.

Line 282: "It has to be taken into consideration that" → "It should be borne in mind that"

Done.

Line 284: "On the contrary" → "In contrast"

Done.

Line 285: "less" → "fewer", or "a narrower range of" if that's a better description.

We use 'fewer'.

Line 437: "exemplification" → "example"

Done.

Replies to Referee #2 on the manuscript 'Assessment of the error budget for stratospheric ozone profiles retrieved from OMPS limb-scatter measurements ' by C. Arosio et al.

We thank the reviewer for the time she/he spent reading the manuscript and constructively commenting on the paper. In the text below, we address the comments from the Referee #2. Referee's comments are shown in italic and authors' responses are highlighted in blue.

Dear authors, Congratulations on this comprehensive analysis and well-written paper. It was a pleasure to review this work. Error/uncertainty characterization of satellite retrievals is more and more gaining the attention it deserves, and this paper clearly demonstrates the way to go about this. It is (almost, see below) complete in terms of identified sources of errors, and the propagation using numerical techniques ensures sensitivity to those tricky situations where random errors turn into systematic contributions.
The only source of errors I did not see addressed is that related to horizontal smoothing: if the actual sensitivity is not centered on the tangent point (and the atmosphere is inhomogeneous) this will lead to additional errors/uncertainty. Addressing this is not trivial (you'd need horizontal AKs and a 3-D model of the atmosphere, e.g. from a reanalysis), but it has been done before, e.g. for MIPAS on ENVISAT(von Clarmann et al., AMT, 2009, and application in Cortesi et al., ACP, 2007). I think this deserves to be at least mentioned.
That said, my other comments are all minor. See below for the list.
Again, congratulations for this work!
A Referee.

Thank you for the appreciation and the positive review.
It is a good point about the horizontal smoothing and uncertainties coming from the inhomogeneity along the line of sight, we agree that it is also a relevant source of error especially in the presence of sharp gradients along the line of sight. Unfortunately, we are not able to investigate this issue properly because our radiative transfer model and retrieval approach are 1-D. To address your comment we added to the introduction some text about this topic, as follows:
'Another source of error which is not considered in this study but relevant for limb observations, is related to the horizontal smoothing, tackled for example in von Clarman et al. (2009) and Cortesi et al. (2007). Namely, limb observations smooth the atmospheric variability over a 250–400 km region around the tangent point (TP) along the line of sight (Zawada et al. 2008). In addition, a 1D retrieval cannot account for gradients along the line of sight, leading to an additional uncertainty component, which is not expected to be relevant on average, but might play a role for atmospheric scenes characterized by strong gradients along the line of sight, e.g. in the presence of the ozone hole (Zawada et al. 2008) or sharp reflectivity gradients.'

Abstract:
- Maybe add a few words on the impact of the cross section source?

We added a mention of this in the abstract, although we don't consider this contribution in the final error budget.

Intro:
- I'd reorder the arguments for the importance of the error characterization: foremost, correct uncertainty estimates are needed to ensure correct geophysical interpretation. Only in 2nd order are they important in the validation. Actually, the validation should validate also these

ex-ante/prognostic uncertainty estimates (as you do).

Thanks, we re-formulated this paragraph accordingly.

- Can you not create a short reference, e.g. Rault and Loughman (2012), hereafter RL12. And then change "The authors" into "RL12". To be checked against the AMTD style guidelines.

I have checked in the guidelines but could not find any information about such abbreviations, I will ask in the typesetting phase if it is possible.

Sect. 2:
- I guess there is also an official NASA retrieval that deserves to be mentioned before presenting the Bremen retrieval? Or is the Bremen retrieval the official product?

Thanks for pointing this out, we added a mention of the official NASA product at the beginning of the section.

- The SNR is in principle something to characterize random noise. But perhaps you

The reviewer is right, there are some systematic structures in the fit residuals, which are not explicitly taken into account. This however seems to have a negligible impact on the retrieval, as depicted in Fig.2 below (see discussion there).

- out of curiosity: how does your vertical resolution estimate compare to actually measuring the FWHM of the AK row? No need to invest much time in this, as you wish.

In Fig.1 below you can find a plot of the vertical resolution together with the FWHM of the AK, for an example case. As you can see they agree very well.

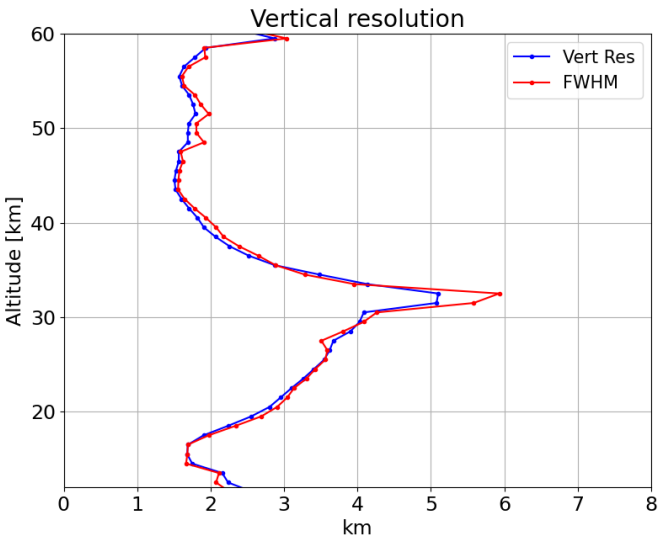

Figure 1: Comparison of vertical resolution with AK FWHM for an example case.

Sect. 3:
- if I understand correctly, you first derive a SNR from the residuals and then use this as starting point for the Gaussian generator, forcing the input errors to have a normal distribution.

Could you not take the actual residuals (of many retrievals) as a distribution to draw (randomly) from? That way, you'd also propagate potential systematic effects, no?

Thanks for this comment, which is directly related to the previous comment about systematic errors and SNR. Yes, we derive the SNR from the residuals and generate 50 normally distributed samples. We then compared the results of the usage of these 50 generated noise samples with the usage of the actual SNR sample from the measurement. In Fig. 2 below we show the results of this comparison, for a case of temperature perturbation (2K). The left panel shows the 50 retrieved profiles, using as inputs synthetic intensities with Gaussian SNR added, together with their mean in black, the average unperturbed profile in blue (same methodology) and a profile retrieved from synthetic intensity but using the actual SNR from the OMPS observation in green. On the right panel the relative differences 'black - blue' and 'green - blue' are shown. After looking at several cases, we found a very good agreement. This points out that possible systematic errors in the SNR are a secondary effect.

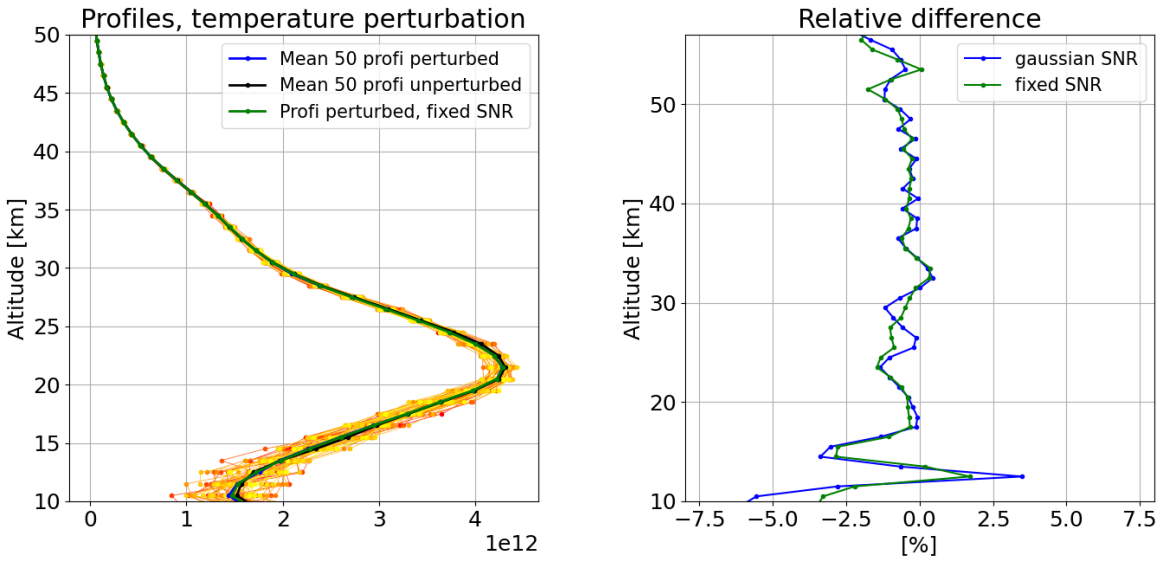

Figure 2: On the left panel, 50 synthetic perturbed retrievals with their mean (black) and the average unperturbed profiles (blue), when using Gaussian SNR samples. Over-plotted, a single perturbed profile retrieved using OMPS SNR. On the right, relative difference of the perturbed profiles (black and green) w.r.t. to the mean unperturbed, for the two cases.

Sect. 4:

- errors on these parameters can all be assumed to be random?

- Yes, to our knowledge the error on this parameters are essentially random. Only TH error could have systematic biases which are however not know (the known bias are accounted for in the pointing corrections from NASA)

- pointing errors: out of curiosity: is there any drift in the platform that would lead to a systematic but slowly changing pointing error that could cause a long-term drift in the retrieved profiles?

This is a very good question and the NASA team has been investigation pointing issues for quite some time and could not point out a clear drift in the pointing, although an introduced

jump-like correction may have enhanced the drift. An improved L1 data version is being processed by NASA and the possible the drift of the new time series will be checked.

Sect. 5:
- line 278: by "changing the used cross section", I guess you mean the source of the cross sections? For clarity, I'd rephrase to "cross section source" or "cross section database". This also needs to be changed in line 280.

Thanks, yes, in this way it is more clear.

Sect. 7:
- Eq. 11: you're entirely confident that they are all independent? Intuitively, I would agree. But maybe this assumption needs to be made explicit. EDIT: I guess Fig 13 demonstrates that they are not independent. What does that mean for Eq. 11?

Thanks for this comment, it is a point that can be improved. Fig. 13 of the manuscript shows that the parameter uncertainties are correlated, although it doesn't directly imply their interdependence. It is however sensible to assume some interdependence between them, e.g. increase in aerosol extinction impact also the reflectivity. However, since we have no relevant information regarding their dependence we have to assume them independent. A comment to Eq. 11 has been added.

Typos:
- abstract, line 3: complaint → compliant
- line 36 (Sheese et al. ref): \cite → \citep
- line 70: increases → decreases; up to → down to
- line 92: parameters → parameter

All the typos have been addressed.